Michael Wornow and Theo Walker 5/1/18

US 34: Final Project Write-Up

Civil War Data Analysis Project

View the live project here: <https://us34-final-project.herokuapp.com/>

View the code behind everything here: <https://github.com/Miking98/us34-final-project>

For our final project, we hoped to substantiate themes from the course (and hopefully find new interesting patterns as well) by running computational analyses of Civil War datasets. After surveying several categories of data and uncovering a few interesting trends, we created a website as the end product through which we could present our findings in a user-friendly way and allow users themselves to sift through the data and come up with their own unique insights. As Professor Staffer has argued, during the Civil War era the emerging technology of photography was the democratic medium of choice for many, including Frederick Douglass, due to its ability to convey the unfiltered truth of reality. Connecting this to the present day, we believe that data visualization technologies powered by sophisticated algorithms will be the tool of choice for enabling democracy in the 21st century, as the tools that we developed (and now display on our website) allow anyone and everyone to look directly at the data and uncover the truth behind it. Just as photography gave Americans in the 19th century an unflinchingly realistic view of the world for the first time in human history, so too will data visualization tools allow the general public to (for the first time ever) directly analyze and interpret the data that drives policies and high-impact decision-making in society.

We split the project into four main stages: 1) Finding accessible and interesting datasets, 2) Scraping, storing, and compressing those large datasets into a standardized, machine-readable format, 3) Building analysis and data visualization tools to glean insights from the data, and 4) Building a website to publish the tools we developed and allow the general public to access our data. Steps (1) and (2) together took roughly 15 hours total, with a major bottleneck being the planning and development of efficient data structures for storing potentially gigabytes of data in a way that could be quickly queried. Step (3) took roughly 25 hours total, with 10 hours spent learning how to create data visualizations in D3.js and process data using Python's NLTK and Pandas libraries, and then an additional 15 hours to actually design, build, and debug the visualizations present on the website. Finally, bringing everything together into a centralized website took about 10 hours total of coding and setting up a server to host our files (which also added time to Step (2) since many of our files were initially too large to host for free until we managed to compress them more efficiently). Thus, not counting time spent writing up the various sections of the website and actually identifying interesting trends in our data, just building the various tools seen on our website took both of us over 50 hours. We were very happy with the way the website and project turned out overall, and thus were extremely happy to see our efforts pay off in the end.

The first major way in which we succeeded in meeting our initial goals of investigating key themes of the course was through our creation of our Letter Analysis tool. This data visualization tool allows a user to enter a phrase of one or two words and then see how the frequency of that phrase's usage in letters written during the Civil War varied with respect to time. While we initially had somewhat ambivalent hopes that this aspect of the project would yield interesting results, we are incredibly happy to report that it met and exceeded all of our wildest expectations. The tool illuminates all sorts of fascinating trends, from the changing spiritualism during the war as theorized by Faust to the rise and fall of army commanders to discussions surrounding important policies, key battles, and their aftermath. It can even be used to substantiate or refute modern positions, such as the common claim that the Civil War was fought over “state’s rights” -- this phrase appears in literally none of the private correspondence we sampled, illustrating how people of the time truly understood that the Civil War was a battle over slavery. We have included a list of a few interesting trends that we uncovered on the website so that any users can investigate these instances for themselves. Ultimately, the Letters analysis tool shows users how although communication during the Civil War era was much more restricted and isolated than it is in today’s hyper-connected technological environment, public and private discussion still followed well-defined movements depending on current events and updates on the status of the war.

Another way in which we were able to illuminate historical data was through our exploration of data on white Union Army Recruits. Based on Census data on 39,000 white Union army veterans, we created an interactive map that allows the user to explore recruitment trends on a county-to-county basis. This data set was by far more coherent than any we could find for Confederate soldiers or for black soldiers fighting for the Union, but we hope that these data points could eventually be added to our work. The map reveals pockets of local wealth along the East Coast and, somewhat more surprisingly, in the Westmost portions of the Union, potentially due to the value of the land holdings for settlers out West. Furthermore, the map reveals pockets of Union recruits who lived in Confederate/border-state counties but chose to enlist in the Union army. We found these soldiers to be older and poorer, on average, than those who enlisted from within the Union.

Our final section within the website is the Modern Trends map, which outlines connections between states' allegiances during the Civil War and attitudes towards more modern policies. For example, we compare the Union-Confederacy map, states with education segregation laws in 1954 pre-Brown v. Board of Education, and states which voted against the 1964 Civil Rights Act. These data selections allow for us to directly view the manner in which the Civil War is still being culturally and politically fought to this day. The timeline of the map culminates in a final data sample in which modern rates of felon disenfranchisement among African-American adults are examined, and we see that rates of African-American disenfranchisement maps closely to the allegiance of a state during the Civil War/Jim Crow segregation policies in the early 20th century. Interestingly, convicting someone of a felony under harsh sentencing laws is the only form in which modern day "slavery" or, to be more technical, "servitude" can legally exist in America, for the 13th amendment (ratified after the Civil War during Reconstruction) explicitly carves out this modern-guise for slavery: “Neither slavery nor involuntary servitude, except as a punishment for crime whereof the party shall have been duly convicted, shall exist within the United States, or any place subject to their jurisdiction." Thus, it seems, the negative impacts of the unjust seizure of a person's labor are still being felt in modern American despite the Civil War ending centuries ago.

In terms of the data analysis aspect of the project itself, we faced many difficulties throughout the duration of the project, both in terms of technological bugs and sparse/incomplete/poorly formatted datasets. Through a bit of clever engineering and the compilation of many separate datasets, however, we were thankfully able to overcome most of these issues. Unfortunately, we had one major goal which we were not able to accomplish, mostly due to constraints beyond our control. We initially intended to use the Twitter API to perform a sentiment analysis on discussions surrounding the Civil War, which would provide a window into the ways that the war’s legacy is still being interpreted. However, after initially formalizing our application to register as developers with Twitter, we did not receive a response. This area remains as one way in which we could expand this project and find even more trends in our society to which we could draw connections from our other analyses of the Civil War period.

We hope that through our data analysis project, and the resulting website platform, users can both garner historical insight into media trends in the Civil War period and see the ways in which the war is still being fought to this day. This project was very illuminating for us personally, and we hope that the visitors to our site will both feel the same way and create historical data projects of their own!